



UNIVERSITY OF CALIFORNIA PRESS
JOURNALS + DIGITAL PUBLISHING

The Bayesian Way to Relate Rhythm Perception and Production

Author(s): Makiko Sadakata, Peter Desain and Henkjan Honing

Reviewed work(s):

Source: *Music Perception: An Interdisciplinary Journal*, Vol. 23, No. 3 (February 2006), pp. 269-288

Published by: [University of California Press](#)

Stable URL: <http://www.jstor.org/stable/10.1525/mp.2006.23.3.269>

Accessed: 09/07/2012 10:33

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



University of California Press is collaborating with JSTOR to digitize, preserve and extend access to *Music Perception: An Interdisciplinary Journal*.

<http://www.jstor.org>

THE BAYESIAN WAY TO RELATE RHYTHM PERCEPTION AND PRODUCTION

MAKIKO SADAKATA

*Nijmegen Institute for Cognition and Information
(NICI), Radboud University of Nijmegen*

PETER DESAIN

*Nijmegen Institute for Cognition and Information
(NICI), Radboud University of Nijmegen*

HENKJAN HONING

*Nijmegen Institute for Cognition and Information
(NICI), Radboud University of Nijmegen; Institute for
Logic, Language & Computation (ILLC) / Music
Department, University of Amsterdam*

MEASUREMENTS OF THE PERCEPTION and production of simple rhythmic patterns have been shown not to be in line in some cases. In this study it is demonstrated that a Bayesian approach provides a new way of understanding this difference, by formalizing the perceptual competition between mental representations and assuming possible nonuniform a priori probabilities of the rhythmic categories. Thus we can relate the two kinds of information and predict perception data from production data. In this approach, the contrast between rhythm perception and production data, taken from different studies in the literature, was shown almost to disappear, assembling independent prior probabilities from counts of patterns in corpora of musical scores, or from a theoretical measure of rhythmic complexity. The success of this Bayesian formalization may be interpreted as an optimal adaptation of our perceptual system to the environment in which the produced rhythms occur.

Received April 26, 2004, accepted February 17, 2005

Temporal Patterns

PROCESSING SEQUENCES of short time intervals plays an important role in our everyday life, for instance, in picking up stress patterns in

conversation and in experiencing music. The importance of studying time relations as a mental phenomenon had already been brought up by the end of the nineteenth century (Jastrow, 1890). Since then, perception of time and action in time have attracted much empirical work (e.g., Fraisse, 1984).

Sequences of time points, marked by events, that is, clicks or onsets of notes, are the domain of these studies, though they are usually specified as a sequence of time intervals between events (inter-onset intervals). In musical scores, the notation of time intervals that constitute a rhythm is based on simple integer relations, and rhythm can indeed be represented as a sequence of integers. The term *rhythm* in this study will be used to mean such symbolic sequences. However, deviations from these perfect ratios in the performance of a musical score are usually large and cannot be interpreted entirely as noise. They partially constitute intended timing patterns that can communicate the structure of the piece (Sloboda, 1985). In this study, the term *performance* will mean a sequence of real-time intervals that carries both the rhythm and the expressive deviations.

Humans have a highly developed cognitive system for processing these sequences. The complexity of the mechanism stems from the fact that the two domains of information interact: a symbolic representation for coding rhythmic structure and a way to represent the small continuous deviations that make up the expressive performance. Note that the same rhythmic sequence can be played with various kinds of expression; for example, it can be made to sound swinging or laid-back by introducing small deviations from strict mechanical timing. Thus a notion of best, perfect, or ideal performance of a rhythm can never exist: it depends on the chosen style and the interpretation. Although a symbolic discrete code and continuous information are communicated when a rhythm is performed and subsequently perceived, both types of information become indistinguishable by being combined before they are transmitted as a sequence of time points through the same one-dimensional channel. Thus a large deviation in timing may very well upset the perception of the rhythmic structure itself.

There has been some evidence for categorical perception of rhythm. This process of perceiving the rhythmic structure for a performance is characterized by an increased sensitivity for detecting performance differences near the boundaries of the categories. Clarke (1987) conducted these experiments and showed the existence of categorical boundaries between specific rhythmic patterns. He also demonstrated that metric context (triple vs. duple) causes a shift in the position of the boundary. Schulze (1989) examined rhythmic categorization by using a different experimental setup, including varied tempi. He observed that subjects were able to identify the rhythms reasonably well, in spite of tempo variation. Desain and Honing (2003) specified the systematic mapping of continuous time intervals to rhythms for three-interval patterns and showed that the way categories are formed is affected by metric context. But in all these studies it is quite obvious that, while one perceives the rhythmic structure of a performance, the continuous information is still accessible, allowing one to perceive the expressive character of the performance.

In studies of music performance and expressive timing, it has been shown that there is no neutral, inexpressive way in which only the symbolic structural part of a rhythm can be communicated. Besides, expressive timing is not a random deviation from mechanical performance but has a certain regularity. In general, systematic deviations are usually observed (e.g., Gabrielsson, 1999, 2003) to be linked to the structural units in the piece (bars and beats, phrases, voices) (Clarke, 1985; Palmer, 1997; Sloboda, 1985). Several studies showed that playing an impassive performance, without any expressive deviation, is not even possible (Palmer, 1989). Repp (1992, 1995, 1999) has shown that deviations from a mechanical performance in accordance with expected regularities are harder to detect. These findings suggest that expressive timing is obligatory, inherent in the musical performance in a systematic way, and that our cognitive system seems to require it.

The Relation Between Rhythm Perception and Production

There have been many production studies in which rhythm has been characterized as expressive renditions of sequences of integers, either with strictly controlled experimental material (Gabrielsson, 1974), for full music performances (Timmers, 2002), or somewhere in between (Repp, Windsor, & Desain, 2002). The perceptual topic of the distribution of performance timing that

allows for perception of a specific rhythmic structure has received less attention but has been investigated as well (Clarke, 1987; Desain & Honing, 2003; Schulze, 1989). Nonzero mean-time deviations from strict mechanical timing are commonly reported. It is surprising, however, that the reported means of the deviations from strict mechanical timing are often not consistent between perception and production studies. Consistency of perception and production would be an obvious assumption if we communicated with others and listened to ourselves while producing rhythm. Much classical work on the processing of rhythm perception and production has been based on this assumption (e.g., Eisler, 1976), but studies often focus on only one of the two processes, which might explain why the inconsistency has been long overlooked. In the past 20 years, however, there have been more studies in which the two processes were studied in conjunction (Drake, 1993a; Povel, 1981; Repp, 1992, 1995, 1998; Sternberg, Knoll, & Zukofsky, 1982), and they report that observed values of rhythm perception and production are not always consistent. For example, Sternberg et al. (1982) found that the durational ratios of perceived two-interval rhythms (using a perceptual judgment task) and those of produced rhythms are different, especially for short intervals (an example of discrepancy is presented in Figure 6). The perceptual deviations were found toward enhancing the contrast between two intervals while the production tendency was toward assimilation of two intervals. Thus, a perceived rhythmic category seems to occupy a separate region of the space of all possible performances as its performed counterpart. Taken at face value, this curious fact constitutes counterevidence for theories that postulate perception and production processes as closely integrated.

Sternberg et al. (1982) proposed a model in which rhythm perception and production tasks share a common analog representation but contain several internal transformations of the temporal patterns. The model does not require that the two tasks share these transformations, which accounts for the discrepancy. There have been, however, more claims associating characteristics of perception and production in understanding the cause of this discrepancy. For instance, some authors postulate, as reason for a deviation from the mechanical timing in rhythm production, that we compensate for peculiarities of our perceptual system; we might compensate for a perceptual tendency to hear intervals short by playing them longer (Drake, 1993a; Ihre, 1992; Penel & Drake, 1998, 1999). Others claim the reverse: perception is constrained by production. For example, the learning of musical production evokes

musical expectation that interacts with the way the listeners perceive temporal patterns (Repp, 1992). Yet another explanation states that perception and production are not in a causal relation, but each interact with the other in relation to the musical structure; both tendencies found in the perception and production are restricted by the musical structure itself (Repp, 1995). Yet none of these theories can adequately explain and predict the differences found so far.

A more fundamental issue may need addressing when rhythm perception and production are compared; it lies in the presence or absence of competition between the mental codes for varied rhythms and the possible nonuniform nature of the competition.

Rhythm Perception and Production Tasks

Rhythm perception and production are quite dissimilar tasks. In rhythm production, one mental representation of a rhythmic code is active. Its repeated realization, via a motor program, yields a distribution around a certain timing pattern. In rhythm perception, the space of possible timing patterns is probed. The stimulus is presented and a rhythmic code must be chosen in response. Several codes may be possible candidates for a certain stimulus. Thus, in a perceptual task the mental representations are in competition, whereas in a production task, the choice of the code to be activated is clear, as it is usually presented in the instruction.

Moreover, in production, rhythms are activated in a simpler way because only the target rhythm is selected and performed. In perception, however, mental codes for rhythms are in competition for selection as a perceived rhythm and as a more stable or simple representation, even if it constitutes a less close fit, or may prevent the choice of a closer but less stable fit. Furthermore, this competition may also be biased on the response side because selecting an unlikely rhythm, one not often heard, may be not an optimal choice. Therefore, certain, commonly occurring rhythms attract more responses (the areas in performance space that represent these rhythmic categories are larger) than others: the competition in perception may well be biased.

The difference in task characteristics can, in a fundamental way, influence the distributions of the empirical data. Consequently, comparing means and variances of perception and production data, as in Sternberg et al. (1982), may not lead to valid conclusions. We will introduce the necessary probabilistic method, Bayesian modeling, to address this issue and check this solution's performance on the empirical data.

Bayesian Modeling

A Bayesian Approach

In the Bayesian approach, the probability of a hypothesis being true, given an actual observation, is derived from the probability of the observation, given the hypothesis is true. In the calculation, the a priori probability of the hypothesis, in the context of all possible hypotheses, is taken into account. A Bayesian approach in perception and cognition was first introduced in signal detection theory, which was developed to investigate optimal strategies for the detection of signals in the presence of noise (Green & Swets, 1966; Tanner & Swets, 1954). Since the late 1970s, the quantitative application of the Bayesian approach has been applied in diverse areas of research.

The hypothesis that biological perceptual systems can be explained by using a Bayesian approach has been tested in the field of visual perception with much success (Knill & Richards, 1996). For example, Bayes rule was used to give precise predictions about the perception of visual movement (Weiss, Simoncelli, & Adelson, 2002), and it provided a basis for the explanation of visual illusions (Geisler & Kersten, 2002).

The power of Bayes rule has been fully exploited in Bayesian inference in more complex domains (these can be formalized by so-called graphical modeling; see Jensen, 2001). It has even been proposed as a general processing method for cognition, modeling upward and downward streams of information (Dayan, Hinton, & Neal, 1995). Often an optimal (perceptual) strategy can be deduced. In our proposal, however, only a simple application of Bayes rule is necessary to relate two conditional probabilities.

In producing a temporal pattern, a rhythm is provided, as symbolic code or musical score, and the conditional probability that a specific performance pattern arises, given this score, is estimated from repeated trials or from responses of a pool of subjects. In perceiving a rhythm, a temporal pattern is presented as performance, and the subject is required to identify the rhythm (the score). The conditional probability that a score is perceived, given this performance, is estimated from the responses. Bayes rule relates these two quantities, formalizing the notion of nonuniform competition. It does so by the multiplication of the production distributions by, possibly nonuniform, a priori likelihoods for the rhythms themselves, followed by a subsequent renormalization. This transformation of production data should, according to Bayes rule, be equal to the perception data, to be shown in detail later. Thus, the

Bayesian relationship highlights both how rhythm perception and production data are the same—as one is derivable from the other—and how they are different, as the observed distributions are transformed versions of each other.

By considering the priors in a purely probabilistic interpretation, the familiarity of rhythms can be estimated by measuring frequency-of-occurrence information, for example, from corpora of musical scores and estimates of the amount of exposure of the subject to these music compositions. We will describe this task in detail later.

Taken in a pragmatic, nonprobabilistic way, priors may be used to reflect something else: some patterns are cognitively simpler or easier to code and memorize than others. Simplicity measure may not be similar to a familiarity estimate, because likelihood of rhythms may be expected to be related to complexity by a bell shape, as composers often avoid both too simple and too complex rhythms, as in visual art, where patterns with a medium complexity are appreciated as more interesting or beautiful (Berlyne, 1971; Birkhoff, 1933; but see Boselie & Leeuwenberg, 1985).

Although it is clear from a strict probabilistic stance that the concept of likelihood is necessary for a correct application of Bayes rule, there remains the question of whether likelihood or simplicity is the most important concept for encoding mental representations (van der Helm, 2000). If simplicity is indeed the central factor in choosing among competing representations, what kind of structure can we expect the set of priors to have, and which temporal patterns can be considered simpler? We will review some of the literature on these issues.

Rhythmic Complexity in Perception and Production

It is well documented that temporal patterns that can be represented as small integer ratios are easier to process than ones needing higher ratios. Spontaneous rhythmic patterns, those produced without an indication of specific rhythmic structure and tempo, are typically made up of only two interval durations whose subsequent ratio is roughly 2:1 (Fraisse, 1946, 1956; see Clarke, 1999, and Fraisse, 1982, for a summary). Also Povel (1981) has shown that the reproduction of two interval patterns (ranging from 1:4 to 4:5) was strongly distorted in the direction of 1:2. The same effect was found in other experiments (Essens, 1986; Essens & Povel, 1985; Summers, Bell, & Burns, 1989; Summers, Hawkins, & Mayers, 1986). Some studies show a preference for duple subdivisions over triple subdivisions (Drake, 1993b) predicting, for example, a higher simplicity for 1:2 than

for 1:3. Furthermore, the difficulty in maintaining a clear distinction of duration patterns has been shown even when expert musicians are forced to produce a complex rhythm at a quick tempo (Peper, Beek, & van Wieringen, 1995; Repp et al., 2002). This result can be predicted by a theoretical account of the complexity of ratios.

One problem in formalizing a notion of rhythmic complexity is the interaction between the rhythmic structure of the pattern (intervallic structure) and its metrical interpretation (hierarchical structure), an aspect often implicit in, or induced by, a temporal pattern. Timing of the production in musical performance usually varied depending on the position in the metric context (e.g., Gabrielsson, Bengtsson, & Gabrielsson, 1983). Many approaches to rhythmic complexity combine information of theoretic and perceptual factors (Pressing, n.d.; Tanguiane, 1993; Shmulevich & Povel, 2000). Alternatively, derived or indirect measures could be considered, including the amount of syncopation. For instance, Longuet-Higgins and Lee's (1984) measure of syncopation strength indicates the amount of syncopation of a rhythmic pattern, given a certain metrical interpretation. A more syncopated pattern could be considered more complex. Nevertheless, because these theories define complexity within a given meter, they cannot be used in our study because the data in the experiments were obtained without control of meter.

There are other hypotheses regarding the complexity of ratios of temporal patterns, which may be formalized as a set of priors in a Bayesian approach. The purely numerical notion of a so-called *Farey tree*, which is sometimes applied to explain the human ability to process different temporal ratios, is a good candidate. For example, Peper et al. (1995) demonstrated transitions of the ratio of different tapping rates realized by both hands at the same time (bimanual tapping ratio) according to this Farey tree. The Farey tree yields a ranking of the complexity of ratios, according to the depth in a tree. The complexities increase as we move from the root (see Figure 1). Here we present hierarchical ratios defined as the duration of first interval divided by the total duration of the pattern (e.g., 1/2 signifies two equal durations, i.e., 1:1, and 3/4 is used instead of 3:1). Note that in this manner the hierarchical durational ratio is always between 0 and 1.

Next to the theoretical notion of the Farey tree, we need to look at familiarity of rhythmic patterns. Although estimating a subject's prior exposure to various rhythms is an impossible task, counting the rhythms in a corpus of musical scores may be taken as a first approximation to likelihood of a rhythm. As the

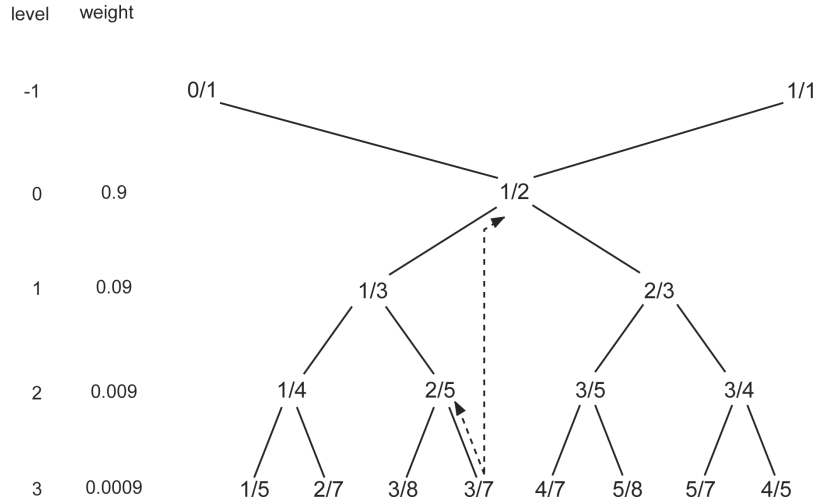


FIG. 1. The Farey theory of the hierarchical ordering of the ratios according to their complexity, visualized as a tree structure. The Farey tree provides a structure of rational numbers, which can be derived algorithmically (Cvitanović, Shraiman, & Söderberg, 1985; González & Piro, 1985).

The ratios at each level (m''/n'') in the tree are obtained from two parent ratios located at a higher level of the tree (m/n and m'/n'), $m''/n'' = (m + m')/(n + n')$. One parent ratio is connected directly to the daughter ratio by a branch (see example arrow). The other parent ratio is found by following the vertical arrow upward until it crosses a branch and then by following the branch upward.

rhythm perception and production data are usually gathered for a fixed number of notes within a repeating time interval (beat or bar), the counting in scores has to account for only n -note patterns that span the unit. Also, the indicated tempo in the score guides the selection of the metric unit under consideration, as it should be roughly the same duration as the unit used in the perception and production experiments, for tempo matters in both rhythm perception and production.

Two Interval Rhythms, Notation and Formalism

For clarity the formalisms used will be based on two interval temporal patterns. Generalization to higher dimensions is straightforward. Let us first characterize a two-interval score rhythm. Assume three successive rhythmic events (note onsets) at score-time z_1 , z_2 , and z_3 , counted in arbitrary units ($z \in \mathbb{N}$). These three points specify two successive note durations ($z_2 - z_1$ and $z_3 - z_2$) and one hierarchical ratio $c = (z_2 - z_1)/(z_3 - z_1)$ of the first note's duration interval regarding the duration of the whole sequence. Each possible rhythm in this domain is thus uniquely identified by a positive rational ratio $c \in \mathbb{Q}$ with $0 < c < 1$, the rhythmic code or category, and we will use this ratio c as the name of a rhythmic structure, irrespective of the notational level (e.g., both the sequence of two quarter notes and of two eighth notes form the ratio $1/2$).

Next define a two-interval performance t . Assume three successive temporal events (onsets) at real-time x_1 ,

x_2 , and x_3 (e.g., in seconds). These three time points specify two successive inter-onset intervals ($x_2 - x_1$ and $x_3 - x_2$) and one hierarchical ratio $t = (x_2 - x_1)/(x_3 - x_1)$ of the first interval with regard to the duration of the whole sequence. Each possible performance is thus uniquely identified by a real ratio $t \in \mathbb{R}$ with $0 < t < 1$. We will use this ratio as a label for a performance.

In a production task, a rhythmic structure c is provided as stimulus or instruction, and a performance t is produced as response. In a perception task, a performance ratio t is presented as stimulus, and a rhythmic ratio c is required as response.

A production data set consists of a number of probability densities over the domain of performance ratios, one for every rhythm c considered, which is illustrated in a schematic way (top panel of Figure 2). Each curve represents the probability for a specific performance t given the instructed rhythm c . Thus, this data set specifies $p(t|c)$,¹ the conditional probability of a performance given a rhythmic instruction. Note that as the curves are densities, the surface area under each equals 1. Because the raw data of a production experiment consist of sets of t collected for each c , the density curves must be estimated from these sets by constructing a histogram or fitting a theoretical continuous distribution.

¹It would be more formally correct to present the conditional probabilities as $p(T|C = c)$, thus a Bayes rule as $P(C = c|T) = p(t|C = c) \times p(C = c)/p(t)$. However, we opt for the shorter notation.

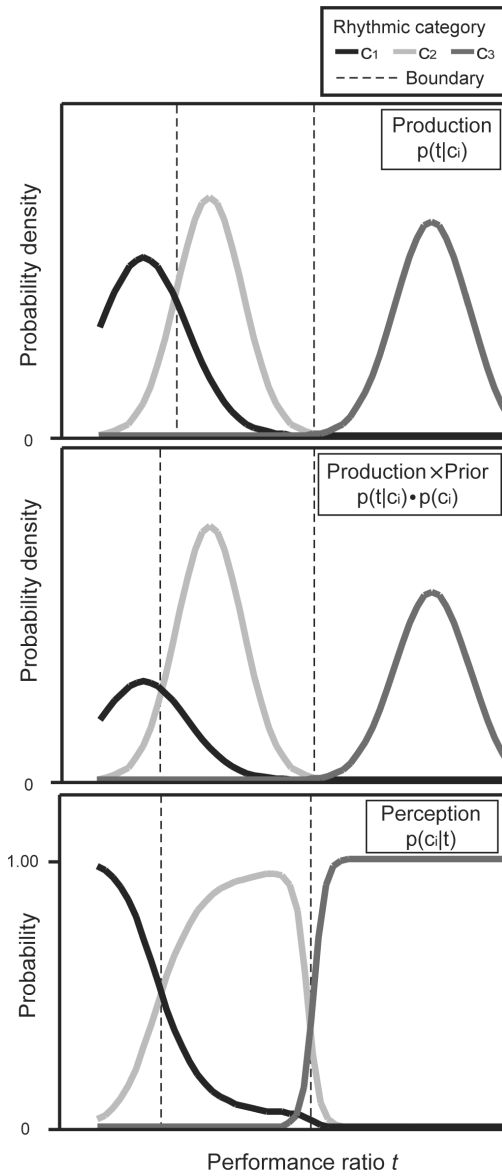


FIG. 2. (a) Example distributions of production data: $p(t|c)$. (b) Example distributions of production data multiplied by priors: $p(c|t) \times p(c)$. (c) Example distributions of perception data: $p(c|t)$ (see formula 1).

A perception data set consists of a number of probability curves over the domain of performance ratios, one for every rhythm c (bottom panel of Figure 2). Each curve represents the probability for a specific response rhythm c , given the presented performance t . These curves are like the receptive fields in visual perception theories, or the tuning curves in the domain of auditory perception. Thus, this data set specifies $p(c|t)$, the conditional probability for a perceived rhythm c given a performance t . Note that as the curves are probabilities,

not densities, here their sum equals 1 for each value of t . Thus, at each t the perception data specify a discrete probability density over the responses. Because the raw data of a perception experiment consist of sets of c collected for each t , the probabilities simply reflect (are estimated by) the response proportions.

A prior data set consists of a set of a priori likelihoods of occurrences, one for each rhythmic ratio c . It is notated as $p(c)$ and reflects the possibly nonuniform exposure to various rhythms.

With these definitions in place, and using Bayes rule, it is possible to define the relation between the constructs. Bayesian modeling provides a framework for reasoning with uncertainty. The notion of conditional probability is central and denoted as $p(a|b)$, which expresses the probability of a occurring when it is given that b occurs. Bayes rule relates the probabilities $p(a|b)$, $p(b|a)$, $p(a)$, and $p(b)$.

Applying it directly to our case, the rule dictates:

$$p(c|t) = \frac{p(t|c) \times p(c)}{p(t)} \quad (1)$$

This can be read as: *the probability of a rhythm being perceived, given a (presented) performance, is equal to the probability of that performance being produced, given that rhythm (as instruction), times the prior probability of that rhythm, divided by the probability of the performance arising in any case.*² The latter term sums over all possible cases (any rhythm). It acts as normalization constant and can be rewritten as

$$p(t) = \sum_i p(t|c_i) \times p(c_i) \quad (2)$$

To return to Figure 2 for an illustration of this calculation, each production density curve $p(t|c_i)$ from the top panel is scaled by a prior probability $p(c_i)$. This yields the middle panel of Figure 2. Then the curves are renormalized, making them sum to 1 for each value of t by dividing by their sum. This maps these likelihoods to the proportions of (forced) responses in the bottom panel, which is taken to predict the perceptual data.

Surprisingly simple, Bayes rule may thus be able to give an explanation for the differences occurring in the means and variances reported for perception and production, as it explains the transformation of the shape

²In this formulation, considering the environment, to-be-perceived and produced performances have been equated. Considering the mental representation of rhythmic structures, task instruction (production) and task responses (perception) are equated as well.

of these curves. Looking at Figure 2, we can see that for each curve (i.e., each rhythm) the performance mean and variance in the two data sets differ, because for perception a strong competing neighbor on one side may skew the response curve.

Before we embark on testing, if the formalism can work on real data sets, there is one caveat. This method can be used only for performance ratios where production data exist ($p(t) > 0$). No prediction can be made for the perception of a performance for which the probability produced is zero for all instructed rhythms. Thus, many rhythms need to be considered in the production experiment. Furthermore, this set should contain all rhythms obtained as responses in the perception task; or vice versa, the possible responses in the perceptual task should be limited to the set of rhythms tested in production. By adding the necessity of an equal tempo in both tasks, these limitations created the difficulty of finding the appropriate data sets for this metastudy.

Hypotheses

Bayesian inference can provide a new way to interpret data, by stating that perception and production are only apparently different, because the difference results from the sensitivity of the rhythmic categories to (nonuniform) competition in perception. Stated in other terms, we hypothesized that perception data predicted from production data using Bayes rule are closer to observed perception data than the production data as a whole.

We will first elaborate our hypotheses in these terms before introducing a more rigorous test. Because of the disparate nature of perception and production data, the statistical test of difference can be carried out using only a rough indication of similarity such as correlation. This goodness-of-fit measure can indicate the closeness of production data to prediction perception data that use various priors. We use a general two-dimensional (rhythm * performance) correlation measure, which is computed from corresponding two variables over all categories and all time points. It gives us the amount of variance in the perception data explained by the production data as well as predicted perception data.

The direct comparison gives us the amount of variance in the perception data directly explained by the production data (r_d^2). Since there are different sets of priors, the prediction using Bayes rule comes in several variants. A first variant poses uniform priors in which all ratios are treated equal. The fit between the perception and this uniform prediction (r_u^2) can be interpreted as an indication of the success of taking into account

only competition. The second option is a nonprobabilistic interpretation using a complexity measure, the Farey tree, giving us r_f^2 . In the next variant, the priors are derived independently from frequency counts in three different corpora of musical scores yielding r_{sA}^2 , r_{sE}^2 , and r_{sT}^2 (Anthem, Essen, and Theme, respectively). First, we expect r_d^2 to be poor when compared with other Bayes predictions. Second, since a uniform prior does not differentiate between rhythmic categories, we assume uniform priors cannot be as good as score priors and Farey priors: $r_u^2 < r_s^2$ and $r_u^2 < r_f^2$. The relation between estimated perception data by score count priors (r_s^2) and Farey priors (r_f^2) is unsure, as the reliability of the estimation of exposure from a corpus of musical scores is not known and neither is the perceptual plausibility of the simple numerical complexity rule. For the final variant, the priors are treated as parameters whose value is found by optimizing the fit between predictions and observations, yielding r_o^2 . This option introduces many parameters, one less than the size of the set of rhythms, and it is obvious that this will result in the best fit: $r_s^2 < r_o^2$ and $r_f^2 < r_o^2$.

For a rigorous test of the significance of the difference between predicted and observed perception data, we applied the Kolmogorov-Smirnov goodness-of-fit test for each performance ratio t . The test examines whether the proportion of probability curves t (4/19 sec, 5/19 sec, etc.) between predicted perception data and observed perception data is significantly different.³ Our hypothesis is that better predictions yield fewer points n at which the predicted probability of responses is still distinguishable from the observed proportion. The raw production data using this test cannot be related to perception. Thus, for a given significance level we predict $n_s > n_o$ and $n_f > n_o$.

The r_o and n_o will be used as an estimate of a ceiling of the method's success: the maximally achievable congruence between a perception and production data set, using only a Bayes rule.

Application of the Method

Material

To yield a relevant comparison across situations with a different experimental method, a careful selection of

³The more common chi-square test cannot be used in this study because there were always categories for which the probability is zero. The underlying variable for the Kolmogorov-Smirnov test is basically required to be continuous, but the known violation of this assumption leads to only slight errors on the conservative side (Hayes & Winkler, 1970).

TABLE 1. Characteristics of the data from the five experiments compared in this study.

Hierarchical and successive interval ratios, and first interval (ms)																
			1/8	1/7	1/6	1/5	1/4	1/3	2/5	1/2	3/5	2/3	3/4	4/5	5/6	7/8
			1:7	1:6	1:5	1:4	1:3	1:2	2:3	1:1	3:2	2:1	3:1	4:1	5:1	7:1
Mode and Data set	N	R	125	143	167	200	250	333	400	500	600	667	750	800	833	875
Perception																
[Desain & Honing]	17	1	166.9 (19.6)	- (-)	169.8 (21.9)	197.7 (22.8)	270 (67.4)	338.6 (60.8)	421 (-)	488.5 (44.1)	579 (-)	636.6 (47.1)	731.2 (47.4)	- (-)	812.5 (26.2)	818.9 (32.1)
[Sternberg (J2)]	3	3-5	59.3 (8.3)	79.7 (10.4)	105.4 (20.4)	154.4 (30.9)	207.3 (50.1)	303.6 (50.8)	- (-)	451.7 (60.7)	- (-)	- (-)	- (-)	- (-)	- (-)	- (-)
Production																
[Repp]	12	6	- (-)	- (-)	- (-)	- (-)	278.6 (27.8)	315.6** (33.6)	364.1* (33.9)	499.6 (17.6)	580.6* (43.5)	631.0** (37.2)	709 (34.3)	- (-)	- (-)	- (-)
[Sadakata]	12	8	- (-)	- (-)	174.3 (23.1)	230.6 (26.8)	267.1 (16.0)	332.5 (12.5)	- (-)	500.9 (8.3)	- (-)	646 (12.9)	721.8 (14.9)	753.2 (21.2)	809.1 (20.8)	- (-)
[Sternberg (P4)]	3	250-1000	156.8 (30.4)	181.4 (30.2)	190.4 (30.0)	- (-)	256.7 (20.3)	- (-)	- (-)	500.1 (20.6)	- (-)	- (-)	743.9 (30.2)	- (-)	814.2 (30.1)	853.7 (40.8)

Note. N = number of subjects, R = number of repetitions. The mean interval duration for the first interval of each ratio is given in ms, and their standard deviations are shown in parentheses. An asterisk (*) indicates an interpolation method was needed; two asterisks (**) indicate the use of an extrapolation method to arrive at the appropriate tempo.

the data sets was needed. The data sets used were collected from the study by Repp et al. (2002, production), Sadakata, Ohgushi, and Desain (2004, production), Desain and Honing (2003, perception), and Sternberg et al. (1982, perception and production), respectively. Detailed data descriptions appear in Appendix 1.

All of the data sets used rhythmic patterns consisting of two intervals whose total duration was 1 second. In the perceptual studies, the subject is presented with a possibly repeated auditory pattern, and the task is to identify a rhythm. In the production studies, subjects were asked to perform a rhythm as a movement pattern by hitting a drum or by playing a piano. (See Table 1 for a list of the rhythmic ratios available in the studies). In Table 1, the means and standard deviations are also listed as calculated from the raw data (i.e., actual responses), or as taken from the original article (in case of the Sternberg et al., 1982, data set⁴). Though many studies show that the actual time durations (tempo) influence musical performance, the issue of time scale cannot be considered systematically in our study, as it proved impossible to obtain access to data sets with more than one tempo condition in common. Thus, although the individual studies may address other tempi, we restricted our analyses to one (moderate) tempo: patterns of two time intervals summing to 1 second. In a few cases, the data were not available at the exact tempo, and a small interpolation was needed.

⁴Sternberg, using moments, translates the “between” measurements into category means and standard deviations (see Appendix C in the original article).

As these data stem from quite separate experimental setups and the procedures and the musical character or naturalness of the tasks was quite diverse, we will first outline the tasks that the subjects had to perform.

In Repp et al., the pianists were involved in a natural musical undertaking: performing monophonic melodies on a piano at a given tempo. The Sadakata et al. experiment was somewhat artificial, by performing in a mechanical way a repeated drum pattern on a pad. For the perception experiments, the free transcription task of Desain and Honing was close to everyday musical activities of musicians and composers. Skilled musicians participated in both the perception and production tasks of Sternberg et al. They identified the rhythmic category of the presented rhythmic pattern by specifying time duration in the perception task and tapped the rhythmic category along the metronome click in the production task.

A Priori Likelihoods: Empirical

To differentiate between rhythmic patterns, frequency counts were derived from databases of musical scores to serve as sets of priors. A very large corpus from a diverse kind of music is necessary to obtain appropriate counts of rhythmic pattern frequencies for the wide range that occur in music. For this, the frequencies of 14 ratios, which occur within metrical subdivisions, were counted from three diverse kinds of music corpora named Anthem, Essen, and Theme (a detailed description of each database appears in Appendix 2). Some ratios, such as 2/5 and 1/7, did not occur at all and

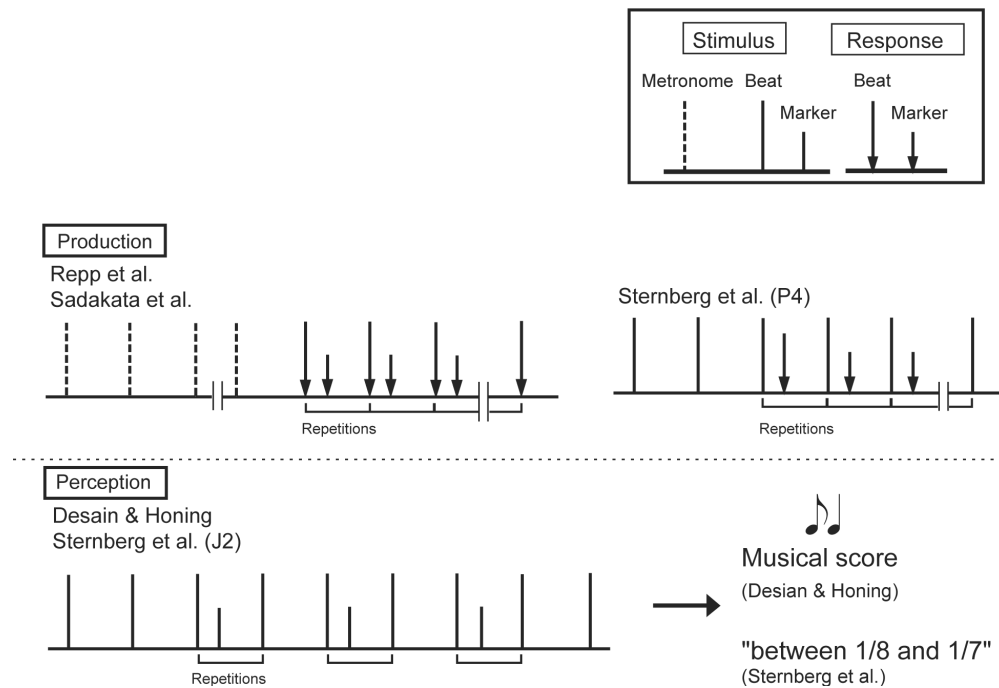


FIG. 3. Example of the stimuli and produced responses in the experiments.

reflect that divisions in 5 and 7 are much less common in Western music (London, 2001).

POCO (Honing, 1990) was used to collect the counts. As the empirical data deal with rhythmic subdivisions of a repeated unit of 1 second, which when presented or performed assume a metrical character, only note pairs that spanned a metric unit (bar, beat, or subbeat) were considered for the counts.

The frequency of occurrence of the ratios used in this study as they appeared in the databases is shown in Table 2. The total number of counted ratios was about 19,000 for the Theme data, 95,000 for Essen, and 4,000 for the Anthems. The range of the counts spans a large range: five orders of magnitude. The frequency of ratios not included in this study (shown as “other” in the table) is very small: .4% for Theme, .002% for Essen, and 0.005% for Anthem.

This result shows that almost all of the relations of two intervals can be classified into the 15 categories used. Table 2 brings out the amount of similarity between the very different corpora. The correlation between counts of two of the databases is always above .78 with a maximum .99.

Complexity Measure as Prior

Though devoid of a probabilistic interpretation, any measure that assigns different weights to rhythms can

TABLE 2. The frequency of all the ratios used in this study as they were extracted from corpora of musical scores: the *essen folksong collection* [essen], the anthem set [anthem], and barlow & morgenstern, *the dictionary of musical themes* [theme].

Ratio name	Essen	Frequency Anthem	Theme
1/8	.000	.000	.001
1/7	.000	.000	.000
1/6	.000	.000	.000
1/5	.000	.000	.000
1/4	.001	.001	.012
1/3	.009	.002	.005
2/5	.000	.000	.000
1/2	.720	.473	.704
3/5	.000	.000	.000
2/3	.122	.012	.081
3/4	.140	.496	.175
4/5	.000	.000	.000
5/6	.006	.001	.008
7/8	.000	.015	.011
Other	.000	.000	.004

Note. .000 should be read as < .0005.

be used as if it were a prior, as long as the measures are positive and sum to 1. In this way, we evaluate the Farey tree, a specific simple ranking of ratios according to their numeric complexity. As the tree (Figure 1) specifies only a ranking, and not a numerical value,

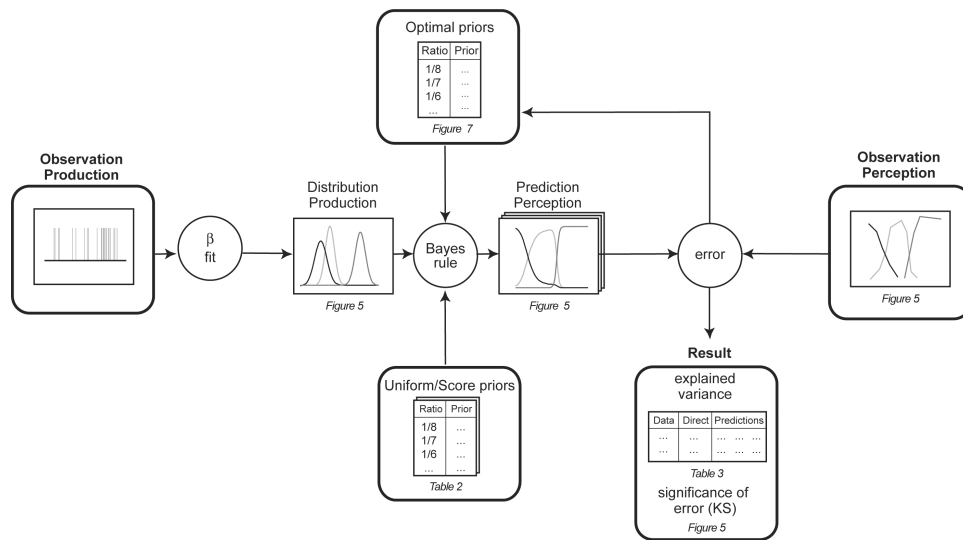


FIG. 4. The paradigm used to compare perception data with production data using Bayesian modeling and the figures and tables in this article.

The inputs are the raw production data (leftmost box), raw perception data (rightmost box), and a set of priors. After fitting a distribution to the production data and applying Bayes rule, using one of the prior data sets, the perception data are predicted. The explained variance of the fit, along with the statistical significance of the remaining difference, is one of the results. The other result, a table of optimal priors, is obtained when the fit is optimized using priors as parameters.

we assigned the root (1/2) the maximum weight and assumed the weight at each level to be a fraction of the weights of the next higher level. We required the weights of the levels used to sum to 1.⁵ This uniquely determines the weights.

Method

Elaborate steps were taken in the computations, partly because the data were not collected with the aim to compare them in the original studies.

Individual observations of *time intervals* are available for each subject, and each ratio was averaged over repeated trials for Repp et al. and Sadakata et al. A schematic of the procedure is presented in Figure 4. The main flow of the information is from left to right. The production observations were modeled by a beta distribution, which can describe these observations quite well, as they are range-limited and usually skewed. (See Appendix 3 for more information about the beta fit.) The fit was done using log-likelihood optimization. The discrete set of probabilities was calculated from the beta distributions using bins around the time grid of the perceptual data as the input for the Bayes calculation. The other input is a set of priors, for which some variants

are available (uniform, three score counts, Farey tree). Bayes rule outputs predicted perceptual judgment distributions, which can be compared with the observed perception data (see the bottom row of Figure 5). The comparison is made by calculating the correlation between distributions, as well as the Kolmogorov-Smirnov goodness-of-fit test. The resulting fit provides the evidence for conclusions about the hypotheses. Furthermore, the mean square error between distributions provides the measure to minimize searching priors that optimally predict the perception data from the production data. This optimization was constrained by requiring the priors to sum to 1.

The processing for the Sternberg et al. data set shares much of the information flow in Figure 4. Unlike Repp et al. and Sadakata et al., only the summary statistics, such as means and standard deviations, are available in this case. Thus, the production distributions had to be reconstructed using a symmetric beta distribution, approximating the given mean and standard deviation. Using the various sets of priors, we predict the perceptual data. However, the distribution of the target, that is, the observed perception data, cannot be reconstructed from the data in the study, as the relative proportions of responses for each ratio category are not available. Thus, we have to resort to deriving the predicted means and standard deviations and comparing them with the observed ones. Minimizing the difference (rms error) between predicted and observed means leads to a set of

⁵The ratios 0/1 and 1/1 were not taken into account, as they do not specify a two-interval pattern.

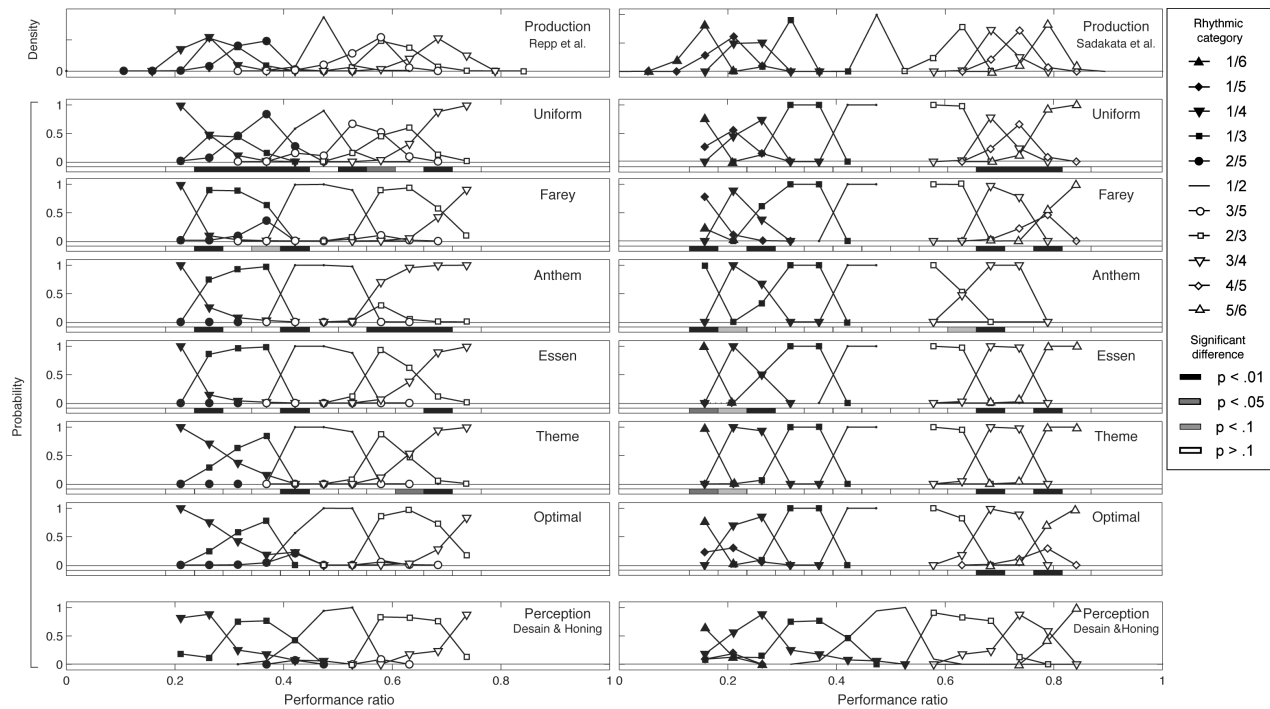


FIG. 5. The observed and predicted distributions of rhythmic categories. On the horizontal axis the duration of the first interval is given; on the vertical axis the probability is represented, either of producing this interval given a rhythmic category or of judging this interval as a proper representation of the given rhythmic category. In the middle rows the perception data as predicted by Bayes rule with different priors are presented.

optimal priors. The results presented by Sternberg et al., using a direct comparison of perception and production statistics, and bypassing Bayes rule, are considered the baseline.

In the case of the Desain and Honing perception study, a few outliers, that is, single responses isolated from the other responses for the same rhythmic category, were observed in the categories 1/6, 1/4, 1/3, and 5/6 (2%). They were treated as errors and excluded from the data. As rhythms used among studies do not completely agree (see Table 1), we selected the rhythms used in each production study for the corresponded perception data to be compared. As a result, a small amount of the Desain and Honing perception data had to be discarded and normalized in the comparison with Sadakata et al. study (3%) and Repp et al. study (6%) respectively, which result in the different shape of the Desain and Honing perception distributions in Figure 5. Thus, the perception data were normalized variously in each case according to the categories used.

From the Repp et al. production study, we could directly use the rhythmic patterns 1/2, 1/4, and 3/4 of the tempo condition normal (total duration is 1000 ms). Yet (linear) interpolation between the “Slow” and “Moderate” condition had to be used for 2/5 and 3/5

patterns, and extrapolation from the “Moderate” and “Slow” condition for 1/3 and 2/3 patterns. From the Sadakata et al. study, all intervals were available at the required tempo. Nevertheless, for both studies the performance tempo was not enforced and drifted slightly over repeated productions of the *time intervals*. As these drifts were very small at this moderate tempo, in the order of 3%, we normalized the *time intervals* to make them sum to exactly 1 second. The preprocessed data were entered into the next stage.

Results

[Repp] and [Sadakata] vs. [Desain & Honing]

Figure 5 shows how, after applying Bayes rule, the distribution predicted from two production data sets approximates perception data. The results using production data from Repp et al. [Repp] are provided in the left column, and the results for Sadakata et al. [Sadakata] in the right column. In both cases, the original production data are shown in the top rows, and the perception data of Desain and Honing in the bottom rows. The second row gives predicted perception data obtained by applying a uniform prior; in the third

TABLE 3. Proportion of the variance in the perception data explained by the production data (and Bayes rule) variance explained.

Data sets		Bayesian model with priors					
Production	Perception	Direct r_d^2	Uniform r_u^2	Farey r_f^2	Score priors r_s^2		Optimal r_o^2
					Anthem	Essen	Theme
[Repp]	[Desain &Honing]	0.34	0.38	0.71	0.47	0.65	0.73
[Sadakata]		0.39	0.66	0.62	0.50	0.68	0.72
							0.74

through sixth rows, the results from various score priors and the Farey prior are presented. The vertical axis shows probability density at the top row and probability for predicted perception and real perception data. The horizontal axis shows the hierarchical ratio on a grid of 1/19, in accordance with the stimulus sampling used in the perception study. Note that a limited range is presented on the x-axis, as perceptual data are available only in that interval. The result of the Kolmogorov-Smirnov goodness-of-fit test is represented as a bar under each prediction. If the prediction on a certain performance ratio is significantly different from perception data, the bar under this time point is gray ($p < .1$, $< .05$) or black ($p < .01$). Nonsignificant difference, which indicates good predictions, is represented as white.

First notice the contrast between top and bottom rows: the nature of the tasks in rhythm production (top) and perception (bottom) is reflected in the divergent curves. It can be easily understood that conflicting means and variances are reported, given that the distributions themselves are so distinct. This dissimilarity is also reflected in the low amount of variance explained by production data (r_d^2), as given in Table 3 under “direct comparison.” Now consider the second row: here are the predicted perceptual data using Bayes rule with uniform priors, assuming all categories equally likely. Accounting for the different nature of the tasks regarding competition produces a considerable change of the shape of the distributions. The success of using uniform priors was varied between the data sets; the uniform priors explained the relation in the case of [Sadakata] quite well, whereas a considerable difference was still observed for [Repp]. The predicted perception data with Farey priors and with score count priors are shown in the next four rows of Figure 5. In the majority of cases, these priors provide fine predictions (see Table 3). The limits of the method are represented in the last row of Figure 5; the priors were considered as parameters and optimized for best fit, significantly raising the proportion of variance explained for [Repp] and [Sadakata].

Although the Farey tree seems counterintuitive in some respects (e.g., 1/4 is intuitively less complex than 2/5), priors from score counts and the Farey tree seem to reflect the relative importance of rhythmic categories to an extent, as they succeed in providing good r^2 s in both production data sets. Yet, it has been shown that the fit can still be much improved at least in the [Repp] set, by optimizing the priors.

Kolmogorov-Smirnov goodness-of-fit tests (see Figure 5) showed that the number of points at which there is a significant difference between predicted and observed perception is considerably decreased by using the optimal priors in both data sets. Thus, the order on the appropriateness of the priors was as expected in [Repp], $r_d^2 < r_u^2 < (r_s^2, r_f^2) < r_o^2$ and $n_u > (n_s, n_f) > n_o$. But in [Sadakata], the order was dissimilar because uniform priors worked well, $r_d^2 < (r_u^2, r_s^2, r_f^2) < r_o^2$ and $(n_s, n_f) > n_u > n_o$.

[Sternberg] vs. [Sternberg]

The row data are necessary for a thorough application of Bayes method; however, using approximations, we can still test if Bayes method works when only statistics (means, standard deviations) are known, for instance, in the study of Sternberg et al.

We explained in the method section that the data distribution of the production experiment (P4) for each ratio was approximated by symmetrical beta distributions. Means and standard deviations of the predicted perception data were calculated, using uniform priors, Farey tree priors, score count priors (Theme, Essen, and Anthem, respectively), and optimal priors. We compared these statistics with the judgment perception experiment (J2). Figure 6 shows the means in the same format as Figure 5 of the Sternberg et al. article. Note that Figure 6 was based on the average response of three participants from the original article, but the response by only one participant was plotted in Figure 5 in the original article. Observed and predicted mean values are plotted against the rhythmic ratio on a log scale.

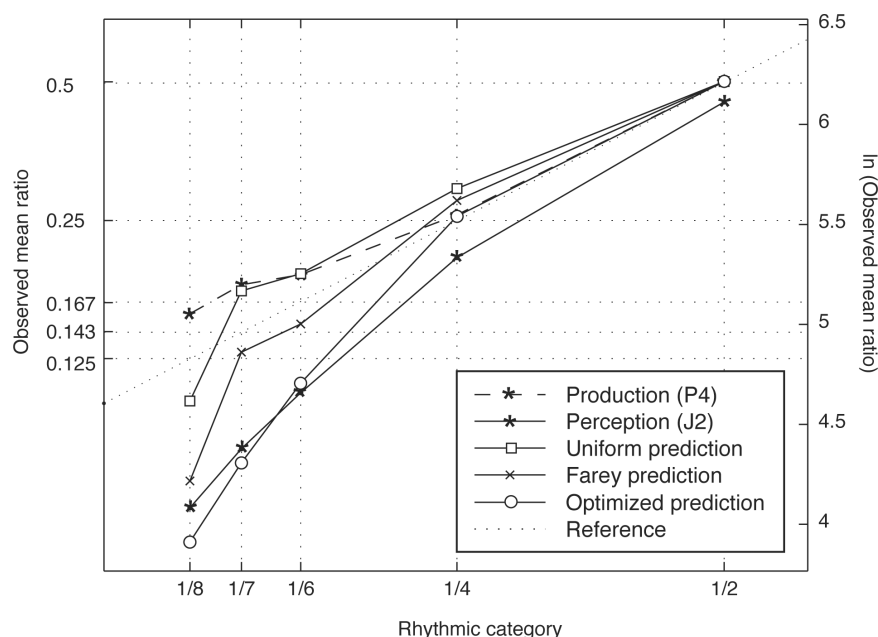


FIG. 6. Observed mean ratios in Sternberg et al. production (P4) and perception (J2) data, and the mean ratios of perception as predicted from the production means by Bayes rule using various priors on a logarithmic scale.

Exact timing provides the reference as the diagonal dotted line in this figure. As reported by Sternberg et al., there is a remarkable discrepancy between the results P4 and J2 (see data marked with * in Figure 6), constituting a contraction of the first interval for the perceptual task and an elongation of the first interval for the production task. This discrepancy becomes especially large for a ratio smaller than 1/4.

In Figure 6 one can see how the Bayesian approach derives predicted perception means from the production data, using different priors. Using the rms error (e) between predicted means and observed means (Sternberg, J2) as criterion, we found similar results as in that of [Repp] and [Sadakata] with $e_o > (e_s, e_f) > e_u > e_d$. The e were smallest when priors were optimized ($e_o = .07$), followed by Essen ($e_{sE} = .10$), Farey ($e_f = .11$), Theme ($e_{sT} = .13$), and Anthem ($e_{sA} = .16$), then Uniform ($e_u = .17$), and Direct comparison ($e_d = .18$). An excellent prediction was made by optimal priors, which makes the distinction between perception and production means almost disappear.

It is important to note that there are few rhythmic ratios contained in the Sternberg study; for example, there is no 1/3 between 1/4 and 1/2 and there is an absence of ratios smaller than 1/8. Note also that results of the score count priors are omitted in Figure 6, as not all predictions could be made due to the lack of these ratios in the score databases.

Priors

Comparing the various priors, reflecting the nonhomogeneity of the space of rhythmic structures, is one point of interest in our study. It is necessary for a good coverage of rhythmic categories to run a good optimization and derive a set of optimal priors. The score counts imply that rhythmic categories included in both the study by Repp et al. and the study by Sadakata et al. cover most of the musically reasonable ratios that can occur within 1000 ms (98.8% of Theme, 100% of Essen, and 98.5% of Anthem). The priors obtained by the combination of production data by [Repp] and [Sadakata] are of interest because they jointly cover these musically reasonable ratios. The data sets of these two studies were combined to arrive at a complete set of optimal priors [Repp-Sadakata].

In Figure 7, the theoretical Farey complexity of ratios, the proportion of occurrence as measured in the score counts, and the optimal priors obtained from [Repp-Sadakata] are presented on a logarithmic scale, to allow for the wide range. In the case of a zero prior, the corresponding value in the logarithmic graph was set to be the lowest rank. At first view their proportions appear to be quite distinct. As the distributions do not extend widely, the size of a prior determines the shape of the curves relative only to its direct neighbors. Thus the optimality of the priors may reflect only the local

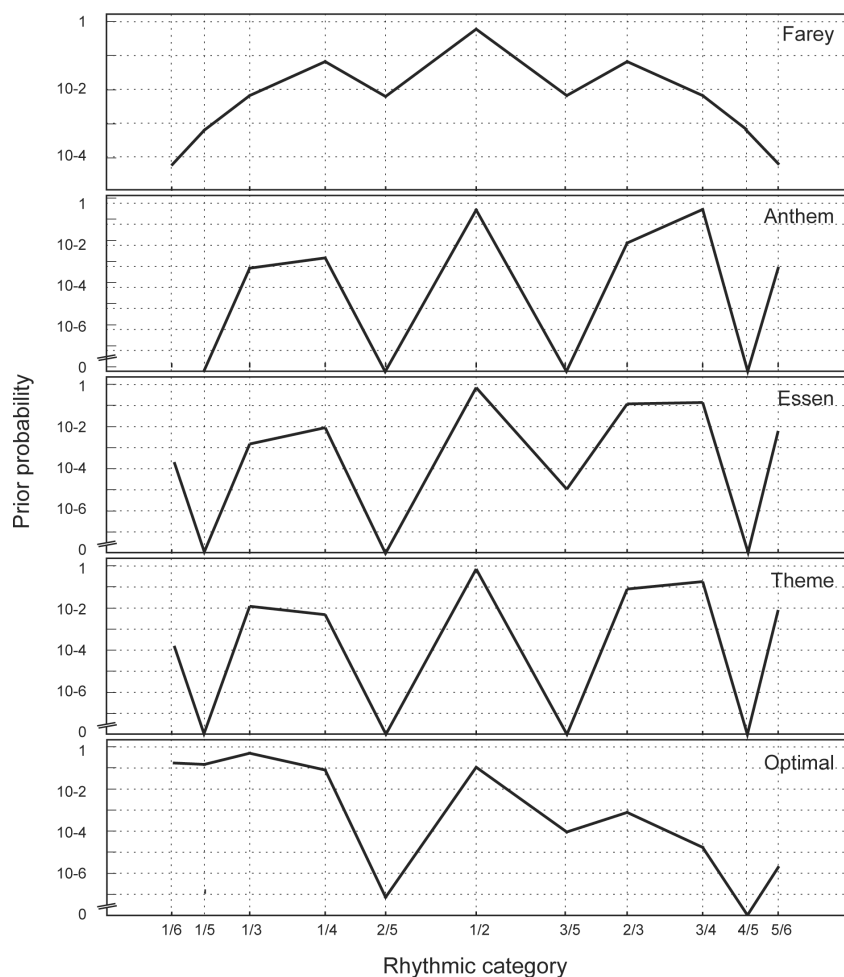


FIG. 7. The optimal priors of the rhythmic categories obtained from Repp-Sadakata, the priors from candidate theories regarding the simplicity of ratios of temporal patterns represented on logarithmic scale and the priors as derived from a Farey tree on logarithmic scale. In the case of a zero prior, the corresponding value in the logarithmic graph was set to be the lowest rank.

relation, and the global structure of the set may not be well expressed, that is, the relative proportion of the leftmost and the rightmost prior may be subject to a much larger estimation error than the relative proportion of two neighboring ones.

Nevertheless, all priors appear to have in common a zigzag pattern. All kinds of priors agree in assigning a smaller value at the patterns having 5 as their denominators; at least they are smaller than the categories right next to them. The Farey tree is number-theoretic and not perceptually inspired, which is, for example, reflected in its 1/5 ratio—less complex than 1/6. The last ratio is usually considered to be perceptually simpler because it decomposes into a hierarchical duple plus triple subdivision. However, the score counts and the Farey tree, when used as priors, are found to give good

predictions and share some characteristics with optimal priors. One observation of interest is well known from other empirical studies: the characteristics that exist in the priors derived from an empirical source (score count and optimal prior) and from a theoretical one (Farey tree). Quintuplets—rhythmic categories with ratios whose denominator is 5—are often mentioned as somewhat unstable patterns to produce or perceive. This instability has been associated with the nature of the mental coding of these temporal patterns (Povel, 1981; Povel & Essens, 1985).

The other artificial nature of the Farey tree is its perfect symmetry, but the assumption that rhythmic patterns retain their characteristics when reversed in time is not realistic. In contrast, Figure 7 shows that the structures of the optimal priors and score priors are

asymmetrical, as optimal priors tend to yield lower priors for long-short than for short-long patterns. The score counts also revealed asymmetrical characteristics but in the opposite direction, as long-short patterns tend to occur more often in musical scores than their short-long reflections. The asymmetry in processing of temporal patterns is also common in empirical studies. For production, Repp et al. (2002) found that 1/3 seems more difficult to perform than 2/3, as the number of trials that participants needed for a good performance according to their own standards is larger for 1/3 than for 2/3. Asymmetry is also commonly found in rhythm perception studies. For example, Desain and Honing (2003) showed asymmetry in the size of permuted rhythmic categories for three interval temporal patterns. Asymmetry in time suggests that auditory perception theories may need to be radically different from visual perception theories in which the symmetry in space plays a strong role. The different gradient between the optimal priors and the empirical score counts found here, however, is a puzzling phenomenon that we do not yet know how to interpret.

Discussion

By using Bayes rule, we have found that rhythm perception data can be predicted accurately from rhythm production data and have concluded that the characteristics of rhythm perception and production processes can be successfully related.⁶

The weakness of the many curve-fitting studies with free parameters—that they do not reveal anything about the flexibility of the theory or the likelihood of other outcomes—has been pointed out by Roberts and Pashler (2000). On this point, it is not surprising that optimized priors yield a good prediction because of the large number of parameters. Indeed, in general (discussed in Desain, Honing, van Thienen, & Windsor, 1998), even a theory that produces a perfect fit to the empirical data is not evidence in itself: there could be alternative explanations that are equally likely.

Nevertheless, the optimized priors tell us what the best obtainable fit is with any corpus, giving a baseline and allowing a comparison with potential competing theories. The real test of the method uses the priors from several candidate theories without any free parameters, such as Farey tree and score counts. It is encouraging for the validity of this approach that

good results were obtained with these priors and that the optimal priors themselves are not very dissimilar to them. Furthermore, their characteristics are in agreement with that found in other empirical studies.

What does this method's success mean in terms of mental processing? Should Bayes rule be considered just a methodological adjustment that makes possible the compensation for the effect of (nonuniform) competition in perception? By defining the strategy a perceiver can use when deciding which category a performance belongs to, we can formulate an answer. The optimal perceptual strategy, the one with the highest expected proportion of correct answers, maximizes the posterior likelihood (in the Bayesian sense) and chooses the rhythm with the highest probability, given the performance. Because human subjects behave close to this strategy, human rhythm perception is optimal, in the sense that it is adapted to, and optimized for, recognition in an environment in which rhythm production takes place—a result that may seem as trivial as it is deep. Progress in perception-action theories may eventually reveal the relevance of this optimality concept. Note that in our approach the production distribution must be fully known to the perceiver, a condition that may seem unrealistic. But advances in machine learning may guide us to a formal understanding of how this knowledge can be learned adaptively.

Implementation of Bayesian concepts for known results in perception and production of rhythm can shed new light and has consequences for theories about music cognition. In an example from the results of Sternberg et al. (1982), the conclusion could be drawn naively: even extraordinary, well-skilled musicians are not able reliably and accurately to produce and recognize time interval ratios in isolation, especially when they are more complex than 1/2. Furthermore, this work reported that rhythm perception and production are different for more complex ratios, and that performed ratios are far from their exact prototypes. The two processes were understood as not completely similar; they share only part of their mechanisms. But in the light of our Bayesian approach, one is drawn to the different conclusion that rhythm perception and production are closely associated; participants behave close to optimal in recognizing temporal patterns, even though the prototypes are far from exact. Furthermore, Bayes rule allows us to attempt to understand the relation between the temporal processing of different patterns while using the concept of priors.

Further questions remain to be answered, such as what the optimal priors suggest and how prior knowledge is acquired. Training subjects on unfamiliar patterns or

⁶It is more difficult to apply the procedure backwards. Given perception data and priors, production data are hard to constrain and predict, especially around the extremes.

comparing rhythm perception, production, and score count data from different musical cultures may be the way to proceed toward a better understanding of the nature of these priors. Again in these cases we expect an optimal attunement of human perception to a world in which human production occurs, even with a changing world. Here individual differences may be modeled as a different set of priors. A more local adaptation may be required when rhythms are presented in the context of a meter or time signature. Because it is known that the “perceptive field,” the area of sensitivity of a rhythmic category, is changed by metric priming (Desain & Honing, 2003), and certain rhythms are more probable in scores with a particular meter (Palmer & Krumhansl, 1990), it might be possible to predict contextual effects by changing the priors, as proposed by Friston (2002). This indicates that our method may eventually have consequences for the difficult cognitive modeling of psychological concepts including priming and attention.

Finally, one important issue not yet discussed is how this approach can be extended to handle more complex rhythms. Surely listeners do not memorize a huge number of distributions for different complex rhythms. Perception of complex rhythm must be based on simple rhythms in a principled way. Both Longuet-Higgins (1987) and Cemgil, Desain, and Kappen (2000) proposed this recursive metric subdivision; however, they assumed categories centered around mechanical timing. As we have shown here, even when one assumes mechanical performances, the perceptual categories may not align with them. A good model of how the perception of more complex rhythms can be derived from distributions of simple perceptual subdivisions is an open and difficult question.

Summary and Conclusion

In this study, we presented evidence that Bayes rule can explain the relation between rhythm perception and production data by assuming that they are identical in a fundamental way. The validity of this approach was demonstrated, and consistent results were obtained under very different experimental conditions and computational setups. Using raw data sets, we simulated the way in which Bayes rule relates the given probability distribution of the production of rhythmic patterns to the probabilities of the perception of rhythmic patterns. Even with limited information of the data set, when only the means and standard deviations are known, it was possible to provide a relevant prediction.

Author Note

This research was funded by Netherlands Organization for Scientific Research (NWO). We are indebted to Bert Kappen and Ali Taylan Cemgil for suggesting the initial idea. We are grateful to Saul Sternberg, Bruno Repp, and Luke Windsor for their consent to use the data and to David Huron and Bret Aarden for providing help with data analysis. John A. Michon, Harold Bekkering, Dirk Vorberg, and Luke Windsor helped to improve the clarity of presentation. Finally, we would like to thank Amanda Brown for her English editing.

Address correspondence to: Makiko Sadakata, Nijmegen Institute for Cognition and Information (NICI), Radboud University Nijmegen, Building B, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands, E-MAIL m.sadakata@nici.ru.nl

References

-
- BARLOW, H., & MORGENSTERN, S. (1948). *A Dictionary of Musical Themes*. New York: Crown Publishers.
- BARLOW, H., & MORGENSTERN, S. (1983). *A Dictionary of Musical Themes* (Rev. ed.). London: Faber & Faber.
- BERLYNE, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.
- BIRKHOFF, G. D. (1933). *Aesthetic measure*. Cambridge, MA: Harvard University Press.
- BOSELIE, F., & LEEUWENBERG, E. (1985). Birkhoff revisited: Beauty as a function of effect and means. *American Journal of Psychology*, 98, 1-39.
- CEMGIL, A. T., DESAIN, P., & KAPPEN, H. J. (2000). Rhythm quantization for transcription. *Computer Music Journal*, 24(2), 60-76.
- CLARKE, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic Press.
- CLARKE, E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabriellson (Ed.), *Action and perception in rhythm and music* (pp. 59-33). Stockholm: Royal Swedish Academy of Music, 55.

- CLARKE, E. F. (1999). Rhythm and timing in music. In D. Deutsch (Ed.), *The Psychology of music* (2nd ed., pp. 473–500). New York: Academic Press.
- CVITANOVIĆ, P., SHRAIMAN, B., & SÖNDERBERG, B. (1985). Scaling laws of mode locking in circle maps. *Physica Scripta*, 32, 263–270.
- DAYAN, P., HINTON, G. E., & NEAL, R. M. (1995). The Helmholtz machine. *Neural Computation*, 7, 889–904.
- DESAIN, P., & HONING, H. (1999). Computational models of beat induction: The rule-based approach. *Journal of New Music Research*, 28, 29–42.
- DESAIN, P., & HONING, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32, 341–365.
- DESAIN, P., HONING, H., VAN THIENEN, H., & WINDSOR, W. L. (1998). Computational modeling of music cognition: Problem or solution? *Music Perception*, 16, 151–166.
- DRAKE, C. (1993a). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society*, 31, 107–110.
- DRAKE, C. (1993b). Reproduction of musical rhythms by children, adult musicians and adult non-musicians. *Perception & Psychophysics*, 53, 25–33.
- EISLER, H. (1976). Experiments on subjective duration 1868–1975: A collection of power function exponents. *Psychological Bulletin*, 83, 1154–1171.
- ESSENS, P. J. (1986). Hierarchical organization of temporal patterns. *Perception & Psychophysics*, 40, 69–73.
- ESSENS, P. J., & POVEL, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, 37, 1–7.
- FRAISSE, P. (1946). Contribution à l'étude du rythme en tant que forme temporelle. *Journal de psychologie normale et pathologique*, 39, 283–304.
- FRAISSE, P. (1956). *Les structures rythmiques*. Louvain: Publications Universitaires de Louvain.
- FRAISSE, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.
- FRAISSE, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, 35, 1–36.
- FRISTON, K. (2002). Functional integration and inference in the brain. *Progress in Neurobiology*, 68, 113–143.
- GABRIELSSON, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology*, 15, 63–72.
- GABRIELSSON, A. (1999). Music performance. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 501–602). New York: Academic Press.
- GABRIELSSON, A. (2003). Music performance research at the millennium. *Psychology of Music*, 31, 221–272.
- GABRIELSSON, A., BENGTSSON, I., & GABRIELSSON, B. (1983). Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology*, 24, 193–213.
- GEISLER, W. S., & KERSTEN, D. (2002). Illusions, perception and Bayes. *Nature Neuroscience*, 5, 508–510.
- GONZÁLEZ, D. L., & PIRO, O. (1985). Symmetric kicked self-oscillators: Iterated maps, strange attractors, and symmetry of the phase locking Farey hierarchy. *Physical Review Letters*, 55, 17–20.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- HAYES, W. L., & WINKLER, R. L. (1970). *Statistics: Probability, inference, and decision*, Vol. II. New York: Holt, Rinehart and Winston.
- HELM, P. A. VAN DER (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, 126, 770–800.
- HONING, H. (1990). POCO: An environment for analyzing, modifying, and generating expression in music. *Proceedings of the International Computer Music Conference* (pp. 364–368). San Francisco: Computer Music Association.
- HURON, D. (1999). Highpoints: A study of melodic peaks by Zohar Eitan. *Music Perception*, 16, 257–264.
- HURON, D. (2002). Music information processing using the Humdrum Toolkit: Concepts, examples, and lessons. *Computer Music Journal*, 26, 15–30.
- IHRE, A. (1992). *Production and perception of rhythm patterns within one beat*. Unpublished master's thesis, Leiden University, The Netherlands.
- JASTROW, J. (1890). *The time-relations of mental phenomena*. New York: N. D. C. Hodges.
- JENSEN, F. V. (2001). *Bayesian networks and decision graphs*. New York: Springer.
- KNILL, D. C., & RICHARDS, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.
- LONDON, J. (2001). Rhythm. In *The new grove dictionary of music and musicians* (Rev. ed., Vol. 21, pp. 277–309). London: Macmillan.
- LONGUET-HIGGINS, H. C. (1987). *Mental processes*. Cambridge, MA: MIT Press.
- LONGUET-HIGGINS, H. C., & LEE, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception*, 1, 424–441.
- PALMER, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 331–346.
- PALMER, C. (1997). Music performance. *Annual Review of Psychology*, 48, 115–138.
- PALMER, C., & KRUMHANS, C. L. (1990). Mental representations of musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 728–741.
- PENEL, A., & DRAKE, C. (1998). Sources of timing variation in music performance: A psychological segmentation model. *Psychological Research*, 61, 12–32.
- PENEL, A., & DRAKE, C. (1999). Seeking “one” expressive timing. In S. W. Yi (Ed.), *Music, mind and science* (pp. 271–297). Seoul: Seoul University Press.

- PEPER, C. E., BEEK, P. J., & VAN WIERINGEN, P. C. W. (1995). Multifrequency coordination in bimanual tapping: Asymmetrical coupling and signs of supercriticality. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1117-1138.
- POVEL, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 3-18.
- POVEL, D. J., & ESSENS, P. (1985). Perception of temporal patterns. *Music Perception*, 2, 411-440.
- PRESSING, J. (n.d.). Cognitive complexity and the structure of musical patterns. Retrieved August 29, 2003, from <http://psy.uq.edu.au/CogPsych/Noetica/OpenForumIssue8/Pressing.html>
- REPP, B. H. (1992). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition*, 44, 241-281.
- REPP, B. H. (1995). Detectability of duration and intensity increments in melody tones: A partial connection between music perception and performance. *Perception & Psychophysics*, 57, 1217-1232.
- REPP, B. H. (1998). Variation on a theme by Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 791-811.
- REPP, B. H. (1999). Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper. *Perception & Psychophysics*, 61, 529-548.
- REPP, B. H., WINDSOR, W. L., & DESAIN, P. (2002). Effects of tempo on the timing of simple musical rhythms. *Music Perception*, 19, 563-591.
- ROBERTS, S., & PASHLER, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107, 358-367.
- SADAKATA, M., & DESAIN, P. (in preparation). Categorization in the perception of two interval rhythmic patterns.
- SADAKATA, M., OHGUSHI, K., & DESAIN, P. (2004). A cross-cultural comparison study of the production of simple rhythmic patterns. *Psychology of Music*, 32, 389-403.
- SCHAFFRATH, H. (1993). Repräsentation einstimmiger Melodien: computerunterstützte Analyse und Musikdatenbanken. In B. Enders & S. Hanheide (Eds.), *Neue Musiktechnologie* (pp. 277-300). Mainz: Schott.
- SCHAFFRATH, H. (1995). The Essen Folksong Collection in the Humdrum Kern Format [Computer database]. D. Huron (Ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities.
- SCHULZE, H. H. (1989). Categorical perception of rhythmic patterns. *Psychological Research*, 51, 10-15.
- SHAW, M., & COLEMAN, H. (1960). *National anthems of the world*. London: Pitman.
- SHMULEVICH, I., & POVEL, D. J. (2000). Complexity measures of musical rhythms. In P. Desain & L. Windsor (Eds.), *Rhythm perception and production* (pp. 239-244). Lisse, NL: Swets & Zeitlinger.
- SLOBODA, J. A. (1985). Expressive skill in two pianists: Metrical communication in real and simulated performances. *Canadian Journal of Psychology*, 39, 273-293.
- STERNBERG, S., KNOLL, R. L., & ZUKOFSKY, P. (1982). Timing by skilled musicians. In D. Deutsch (Ed.), *The psychology of music* (pp. 181-239). New York: Academic Press.
- SUMMERS, J. J., BELL, R., & BURNS, B. D. (1989). Perceptual and motor factors in the imitation of simple temporal patterns. *Psychological Research*, 51, 23-27.
- SUMMERS, J. J., HAWKINS, S. R., & MAYERS, H. (1986). Imitation and production of interval ratios. *Perception & Psychophysics*, 39, 437-444.
- TANGUANE, A. S. (1993). *Artificial perception and music recognition*. Berlin: Springer-Verlag.
- TANNER, W. P., & SWETS, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401-409.
- TIMMERS, R. (2002). *Freedom and constraints in timing and ornamentation: Investigations of music performance*. Maastricht: Shaker Publishing.
- WEISS, Y., SIMONCELLI, E. P., & ADELSON, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5, 598-604.

Appendix 1

Data Description

Desain & Honing (Perception)

In Desain & Honing (2003), a categorization experiment is described in which subjects respond with a rhythmic category that best reflects a three-interval performance pattern, by using a computer interface for common music notation. More details, full results, and a model can be found in Sadakata & Desain (in preparation). In this experiment, 17 skilled musicians participated. Each stimulus pattern was made up of two time intervals on a time grid of 1/19 of 1000 ms, the minimum duration of an interval being three time grid units and the maximum 16 units. This yielded a set of 14 stimulus patterns. The pattern was repeated three times, embedded in a beat, as illustrated in Figure 3. The participants were asked to use notations commonly encountered in their practice. The set of possible response categories, using the computer notation interface, was still extremely large: thousands of ratios can be constructed using a range from whole-note to thirty-second-note durations, using dotting, ties, triplets, etc. The actual responses used only 18 rhythms of the set.

Repp et al. (Production)

In Repp et al. (2002), the task was to perform simple monophonic melodies with the following rhythmic patterns: 1/2, 2/5, 1/3, 1/4, 3/5, 2/3, and 3/4. Twelve pianists participated and performed from a musical score. A maximum of three attempts was permitted without rehearsal, and the version that satisfied the performer was used. The rhythmic patterns were repeated over six bars and performed in four different tempi; a metronome was used before each performance. Averages over repetitions of inter-onset times were used. The responses are indicated as a probability density at the top of the left column in Figure 5.

Sadakata et al. (Production)

The task in Sadakata, Ohgushi, & Desain (2004) was to perform nine kinds of rhythmic patterns: 1/2, 1/3, 1/4, 1/5, 1/6, 2/3, 3/4, 4/5, and 5/6, in three tempo conditions and two playing modes, mechanical and musical. The former was used in this study. Each pattern was performed from a score and repeated 10 times (see Figure 3). Twelve percussionists participated in the experiments. Averages over repetitions of inter-onset times, excluding first and last bar, were used.

Sternberg et al. Experiment J2. (Perception)

In Sternberg et al. (1982), a number of perceptual experiments are presented. Three highly skilled musicians participated. In experiment J2, the subjects on each trial heard five beat clicks spaced 1000 ms apart, with marker clicks following the third and fourth, as shown in Figure 3. They were asked to judge the intervals from the beats to the markers. This beat-marker interval was varied over a range from a minimum of 43 ms to a maximum of 891 ms. They were presented as four different sets of 24 intervals whose spacing varied in the manner of a harmonic series. The participants selected a response from a set of eight categories, which are “less than 1/8 of a beat,” “between 1/8 and 1/7,” . . . , “between 1/3 and 1/2,” and “greater than 1/2.” The eight ordered categories define seven between-category boundaries on a hypothetical response continuum (see n. 4). The estimated means of the psychometric function for each category and its variability were calculated. As the raw data of this study are no longer available, the means and variances were taken from Table 1 and Figure 5 of the original article.

Sternberg et al. Experiment P4. (Production)

On each trial in this experiment by Sternberg et al. (1982), 12 beat clicks were presented. The same musicians served as participants as in perceptual task (J2).

Participants made 10 consecutive finger-tap productions to produce the ratio specified by instruction. The first response was produced after the third beat click, as illustrated in Figure 3. The ratio names used in the experiment were 1/8, 1/7, 1/6, 1/4, 1/2, 3/4, 5/6, and 7/8. Using the average value and standard deviations of these responses, the probability of occurrence of each category on a response continuum was calculated using a normal distribution. Averages were taken from Table 1, and variances were taken from Figure 5 of the original article.

Appendix 2

Data Description of a Database of Musical Scores

Barlow & Morgenstern, Dictionary of Musical Themes [Theme]

The *Dictionary of Musical Themes* (Barlow & Morgenstern, 1948, 1983) is a well-known theme index containing approximately 10,000 themes from the classical music repertoire. Both melody and rhythm are coded for each theme, as well as for its time signature. The collection consists of about 45% duple, 31% triple, and 24% compound meters.

Schaffrath, Essen Folksong Collection [Essen]

The *Essen Folksong Collection* (Schaffrath, 1993, 1995) contains a large sample of European folk songs, collected and encoded in the format of Essen Associative Code. Presently, 6,251 folk songs are available, although the total number of folk songs in the collection has reached 20,000. The metrical structure of the music (as signified by the time signature) is quite varied: 54% duple, 29% triple, and 17% compound meters. The database has been widely used to test a variety of music theories (e.g., Huron, 1999). Though mostly traditional German folk songs, they have simple rhythmic and metric structure and while regionally restricted, some songs are widely known in Europe (Huron, 2002). Thus it could be considered a reasonable sample of childhood exposure to music.

Shaw & Coleman, National Anthems Collection [Anthem]

The *National Anthems Collection* (based on Shaw & Coleman, 1960) is a corpus of the *National Anthems of the World* constructed for the evaluation of beat and meter induction models (Desain & Honing, 1999). The database contains only temporal information (rhythm and meter, no melodic or other information). The set ($N = 105$) consists of around 90% duple (70% is in 4/4 meter) and 10% triple meters.

Appendix 3

Beta Fit

A beta distribution was chosen for the data approximation because of its flexibility, as it can fit skewed distributions. Furthermore, it has no tails extending to infinity. This is an advantage in the next step as minuscule differences in a set of small long tails (as, e.g., a normal distribution would exhibit) may result, after application of Bayes rule, in very large differences from the category center. This may generate noncontiguous categories. The beta distribution has two free parameters, α and β , that characterize the form of curve and two extra parameters (w and m) to rescale (squeeze) and shift the distribution to any mean and width using a linear transformation. In our study, the parameters α , β , w , and m were estimated from the production data set for every rhythmic pattern, using the maximum likelihood method. An example of the fitted beta distribution (and a normal fit to the same observations) appears in Figure 8. The vertical axis shows probability density, and the horizontal axis shows the duration of the first interval.

The production data for each ratio was thus characterized by the four parameters of the shifted beta distribution, yielding a family of curves illustrated in Figure 9, which also shows histograms of the raw data.

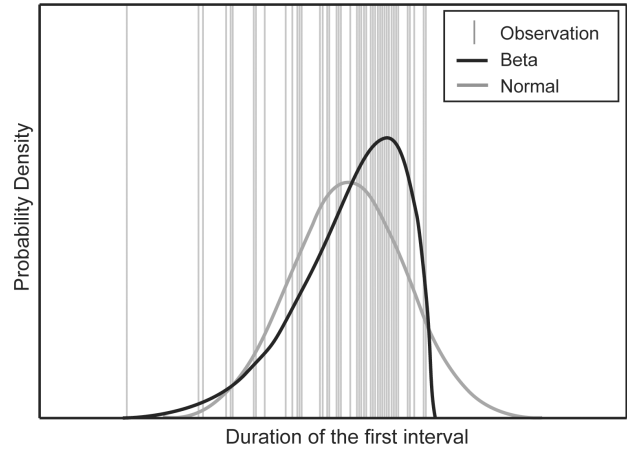


FIG. 8. Example of the relation between observed data and two approximated continuous distributions. Gray vertical lines show the observations, the gray curve represents the data as approximated by the normal distribution, and the black curve represents the data as approximated by a beta distribution.

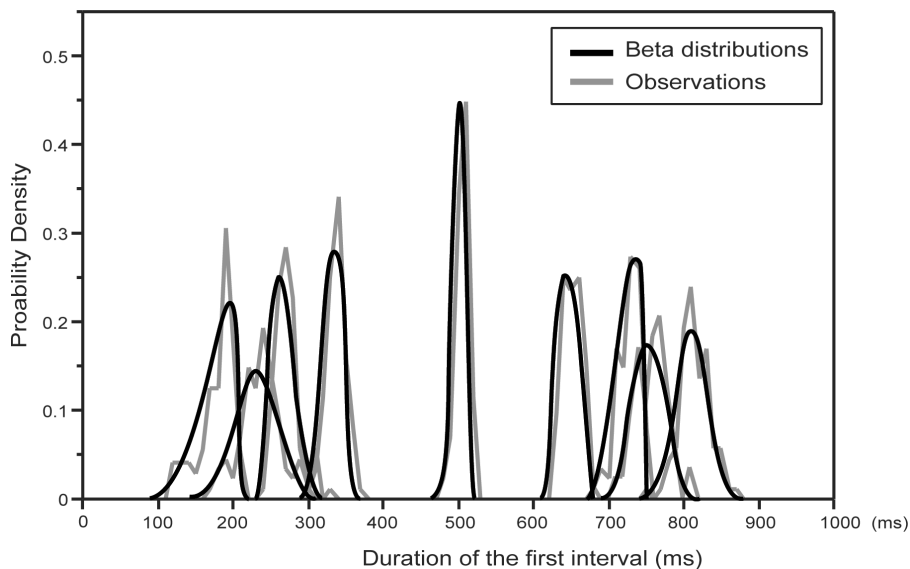


FIG. 9. An illustration of the relation between the histogram of the observed production data (Sadakata et al., gray line) and the approximating beta distributions (black line), for nine different rhythmic ratios.